AD-775 902

CUE UTILIZATION IN A NUMERICAL PREDICTION TASK

Sarah Lichtenstein, et al

Oregon Research Institute

Prepared for:

Office of Naval Research
Advanced Research Projects Agency

February 1974

Unclassified
Security Classification

AD 775 902

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author) Oregon Research Institute Eugene, Oregon	2a. REPORT SECURITY CLASSIFICATION Unclassified 2b. GROUP
3. REPORT TITLE CUE UTILIZATION IN A NUMERICAL PREDICTION TASK	
4. DESCRIPTIVE NOTES (Type of report and inclusive dates) Technical Report	
5. AUTHOR(S) (First name, middle initial, last name) Sarah Lichtenstein, Timothy C. Earle & Paul Slovic	

6. REPORT DATE February, 1974	7a. TOTAL NO. OF PAGES 30	7b. NO. OF REFS 20

8a. CONTRACT OR GRANT NO. N00014-73-C-0438 b. PROJECT NO. NR(197-026) c. ARPA Order No. 2449 d.	9a. ORIGINATOR'S REPORT NUMBER(S) Oregon Research Institute Research Bulletin, Volume 14, No. 3 9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)
10. DISTRIBUTION STATEMENT This document has been approved for public release and sale. Its distribution is unlimited.	
11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY Office of Naval Research Code 455 Arlington, Virginia 22217

13. ABSTRACT

Forty subjects were trained to use scatter plots with regression lines to make numerical predictions of one variable (criterion) from another variable (cue). The subjects were trained on two separate cues, differing in validity. Later, the two cues were presented together, simultaneously for 20 subjects, successively for the rest. Subjects were asked to use both cues to predict the criterion. Instructions emphasized that the two cues were independent of one another, but did not specify how the two should be combined. Initial analyses indicated that a regression model provided an adequate fit to the data, that the subjects showed conservatism similar to the conservatism found in previous Bayesian inference studies. However, more molecular analyses indicated patterns of behavior which consistently deviated from the optimal model. The post hoc hypothesis that subjects were regressing each cue, then averaging the regressed values, was supported by the data for most subjects. Searching for heuristic strategies, rather than relying on the apparent fit of optimal models was advocated for future research.

30

DD FORM 1 NOV 65 1473 REPLACES DD FORM 1473, 1 JAN 64, WHICH IS OBSOLETE FOR ARMY USE.

Security Classification

14. KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
JUDGMENT						
PREDICTION						
DECISION MAKING						
INFORMATION PROCESSING						
UNCERTAINTY						

CUE UTILIZATION IN A NUMERICAL PREDICTION TASK[1]

Sarah Lichtenstein[2]
Oregon Research Institute

Timothy C. Earle
Western Washington State College

Paul Slovic
Oregon Research Institute

How does a decision maker combine information from several sources to arrive at a prediction about the state of the world? How adequate is man as an "intuitive statistician" (Brunswik, 1956; Peterson & Beach, 1967)? Two optimal models have been widely employed in the study of these questions: The Bayesian model, based on probabilities and Bayes' theorem, and the regression model, based on correlational statistics and multiple regression analysis (Slovic & Lichtenstein, 1971). Optimal models typically play a dual role in these studies. First, they serve as standards of ideal performance against which the performance of subjects can be compared. Second, they are used as frameworks for describing the subjects' cognitive processes.

Within the Bayesian paradigm, comparisons of subjects' posterior probability estimates with estimates generated by Bayes' theorem have often shown subjects to be conservative. That is, the human judge fails to extract as much information or certainty from data as does Bayes' theorem (see reviews by DuCharme, 1970; Edwards, 1968; Peterson & Beach, 1967; and Slovic & Lichtenstein, 1971). The finding of conservatism is frequently interpreted as a finding that Bayes' theorem, when modified with suitable subjective parameters to account for the conservative bias, is an appropriate descriptive model of subjects' cognitive processes.

In the regression paradigm, subjects learn a task through repeated trials in which they are presented with several cues, make a quantitative

judgment on the basis of these cues, and are then given the criterion value as feedback. Results of such learning studies have shown that regression equations can provide a good global fit to subjects' judgments based on multiple cues.

These Bayesian and regression studies have led to the conclusion that, since subjects' inferences are influenced by appropriate variables in appropriate directions, normative models provide a good first approximation for a psychological theory of inference (see, for example, Peterson & Beach, 1967, pp. 42-43). However, there has been growing dissatisfaction with this optimistic conclusion. Anderson (1969) has pointed out that global indices can produce high correlations between a model and data even when the model is seriously incorrect. Recent empirical evidence indicates that subjects making probability estimates respond on the basis of task attributes that are irrelevant to the Bayesian model (Beach, Wise, & Barclay, 1970; Pitz, Downing, & Reinhold, 1967; Kahneman & Tversky, 1972; Vlek, 1965). In research with the regression model, Brehmer (1972) and Slovic (1966) have shown that subjects employ different combinatorial strategies for different subsets of stimuli, in violation of the model.

Perhaps the most serious challenge to the view of man as an adequate intuitive statistician comes from the recent studies of Kahneman and Tversky. Kahneman and Tversky (1972) have proposed that when making probability estimates in well-defined, repetitive situations, subjects employ a heuristic called "representativeness," whereby the probability of a sample is determined by the degree to which the sample is similar in essential features to its parent population. They presented several lines of evidence to support their position. First, they found that subjective sampling distributions were insensitive to sample size, a normatively

important but psychologically non-representative factor. Second, they found in a Bayesian task that subjects were strongly influenced by a highly representative sample characteristic irrelevant to the optimal model while being entirely uninfluenced by changes in the relevant, but non-representative, sample statistic.

Kahneman and Tversky (1972, p. 450) note that the principle usefulness of the Bayesian approach to the analysis and modeling of subjective probability depends on whether the model captures the essential determinants of the judgment process. Their research suggests it does not. They conclude, "In his evaluation of evidence, man is apparently not a conservative Bayesian: He is not Bayesian at all."

In a subsequent paper, Kahneman and Tversky (1973) discussed the representativeness heuristic in settings amenable to correlational analysis. They found that people violated the principal assumptions of multivariate prediction methods upon which the regression model is based. For example, subjects' numerical predictions were not properly regressive, and were essentially unaffected by considerations of data reliability.

In the spirit of this recent research, which takes a close, critical look at the validity of algebraic models as representations of cognitive processes, the present study examines performance of subjects in a task conducive to analysis by the regression model. The subjects were asked to predict a criterion number, given two cue numbers which were independent of each other but both correlated with the criterion. Like the typical Bayesian study, but unlike much previous research with regression models, the emphasis here was not on learning, but on the integration of information previously learned. In the training phase, subjects were taught how to use each cue separately, with feedback of the criterion. In the testing phase, subjects were given two cues, one from each cue source, and asked

to combine the cues into a single judgment, without criterion feedback.

The analysis of performance in this task focuses on three questions. First, do the standard parameters and measures of regression models provide an adequate global description of the data? Second, is there an indication that subjects are conservative in their weighting of the two cues, much as they tend to be conservative in the Bayesian analog to this task? Third, does the hypothesis that the normative model is an adequate descriptive model survive a molecular analysis of the data?

With regard to the latter question, use of a regression equation as a descriptive model implies that independent sources of information have an additive impact upon subjects' judgments. Previous research, however, suggests that independent items of information will be averaged, rather than added. Anderson (1973a, b) has demonstrated averaging in a wide variety of contexts, and Kahneman and Tversky (1973) have argued that people predict outcomes that appear most representative of the evidence (which in the present study implies non-optimal averaging).

Performance Measures

Lens Model Measures

The lens model, developed by Brunswik (1952, 1956), provides a method for analyzing judgment within the regression paradigm. This model assumes that certain items of information, called cues, are probabilistically related to a criterion. The multiple regression model is used to express these relationships. The best estimate of the criterion, given the cue values, is

$$\hat{Y}_e = [\sum_i b_{i,e} (X_i - \overline{X}_i)] + \overline{Y} \quad , \tag{1}$$

where $\hat{Y}_e$ represents the predicted criterion value, $b_{i,e}$ represents the regression weights, X_i the cue values, $\overline{X}_i$ the means of the cues, and $\overline{Y}$

the mean criterion value. The subscript, e, stands for the environment, and also serves to remind the reader that Equation 1 represents the model used by the experimenters.

After the subject has been shown a series of cue combinations, and has responded by estimating the correct value of the criterion, Y, for each cue combination, multiple regression techniques can be used to form a linear model of the subject:

$$\hat{Y}_s = [\sum_i b_{i,s} (X_i - \overline{X}_i)] + \overline{Y} \qquad (2)$$

where $\hat{Y}_s$ represents the criterion value predicted from the model of the subject (the subscript, s, stands for subject). The comparison between the model of the environment and the model of the subject forms the central focus of analysis. The details of such analysis have been developed by Hursch, Hammond, and Hursch (1964), Tucker (1964), Naylor and Schenck (1966), and Dudycha and Naylor (1966). This paper uses the notation of Dudycha and Naylor.

Three measures are used as indices of the subject's performance:

r_α: Achievement. The degree of agreement between the criterion values and the subject's responses over n observations. R_α is the correlation between the criterion value, Y_e, and the subject's responses, Y_s.

r_m: Matching. The degree to which the model of the environment matches the model of the subject. r_m is the correlation between $\hat{Y}_e$ and $\hat{Y}_s$.

r_s: Linear consistency. The degree to which the subject consistently utilizes his own strategy as defined by the linear model of his responses. r_s is the correlation between Y_s and $\hat{Y}_s$.

These three indices reflect the general performance of each subject. The indices are interrelated; when the criterion is a linear combination

of the cues (as in this study), the relationship is simple:

$$r_\alpha = r_e r_s r_m \quad ; \tag{3}$$

where r_e is the linear predictability of the criterion, i.e., the correlation between Y_e and $\hat{Y}_e$.

Measures of Conservatism

b-weight ratios. The first index of conservatism used in the present study is based on that proposed by Brehmer and Lindberg (1970). If the subject responds optimally, the b-weights, $b_{i,s}$, which are the slopes of the regression lines relating raw-score cue values and judgments, will be equal to the environmental b-weights, $b_{i,e}$. If the subject systematically treats a cue as if it were less diagnostic than it is, then the b-weight is reduced. Thus, if the ratio, $b_{i,s}/b_{i,e}$, is less than unity, the subject is considered to be conservative in his use of cue i. This measure treats each cue separately; the subject could be conservative in his use of one cue, but not conservative in his use of another.

Revision ratios. The second measure of conservatism here proposed can be used only when the cues are presented sequentially within a single trial and a judgment is made upon receipt of each cue. Optimally, the subject should start each trial knowing that, without any data, the best response is the mean of the criterion distribution. After receiving the first cue, he should revise the mean according to the information contained in the cue. When the second cue is presented, he should revise his first response. The revisions prescribed by the regression model are based on a quadratic loss function.

Assuming that the subject has correctly learned, in the training task, how to respond to one cue alone, the critical judgment is the second one. The optimal revision is the signed difference between $\hat{Y}_e$ given only the first cue, and $\hat{Y}_e$ given both cues. This signed difference, d_e, can be

compared with the signed difference, d_s, between the subject's first and second responses. The ratio of these deviations, d_s/d_e, is here proposed as the measure of conservatism for successive presentation of the cues. If this ratio is positive and less than unity, conservatism is indicated. If this ratio is positive and greater than unity, the subject is assumed to have treated the second cue as more informative than he should have, and counter-conservatism is indicated. When the revision ratio is negative, the subject has responded by revising in the opposite direction from the optimal model; such a response is neither conservative nor counter-conservative.

Method

Stimuli

Two sets of 200 normally distributed two-digit numbers, with mean of 50 and standard deviation of 10, constituted the cue values used for both training and testing. The cue sets were independent of one another. One set correlated .40 with the criterion, while the other set correlated .80 with the criterion. Thus r_e, the linear predictability of the criterion, was $\sqrt{.40^2 + .80^2} = .89$. The stimuli were the same as used by Dudycha and Naylor (1966). The task was presented as a problem of trying to predict a third number, given two "cue" numbers.

Design

Forty male University of Oregon students were assigned to the four groups of a 2 x 2 factorial design. Half of the subjects received, during the testing phase, the Simultaneous presentation of both cues; these subjects made a single response to the two cues. The other 20 Ss received the two cues in Successive presentation during the testing phase, and responded twice, first to one cue, and then to both cues.

The other dimension of the factorial design was order of presentation. Ten subjects in the Successive Group were trained first on the weak (.40) cue, then on the strong (.80) cue; they received the cues in this same order during the testing phase. The other 10 subjects in the Successive Group received the reverse order in both training and testing. Order of presentation for the Simultaneous Group refers only to the order in which subjects were trained on the cues. Due to scheduling difficulties, there were 11 subjects in the .40-.80 Simultaneous Group and 9 subjects in the .80-.40 Simultaneous Group.

Training

All subjects received the same training, except for order of presentation. The subjects were trained in the use of each cue separately, but were never told how to combine the information from the two cues. The training proceeded as follows:

1. After a brief introduction, during which subjects were told the mean values for the cues and criterion, subjects were given 50 training trials. On each trial they were shown a cue, asked to estimate the criterion, and were then shown the criterion value.

2. Subjects were next shown a scatter plot of the cue-criterion relationship for the 50 trials to which they had just responded. They were urged to use the information in the scatter plot on subsequent training trials.

3. Next, subjects were given 10 more trials with feedback.

4. Subjects were then shown a scatter plot of 100 cue-criterion pairs, with a regression line on the plot. They were instructed in the mechanics of using the regression line to make predictions and in the rationale for its use. Finally, they were asked to evaluate the utility

of the regression line for themselves, by using the line to make their predictions on the next series of training trials.

5. Subjects were given 10 more trials with feedback.

6. The subjects were shown a scatter plot, including the regression line, with 150 cue-criterion pairs.

7. Ten more trials with feedback followed.

8. Finally, subjects were given a scatter plot, including the regression line, with 200 cue-criterion pairs.

The same procedure was then followed for training in use of the second cue. After training, all subjects were allowed to retain the final scatter plots for each cue, each containing 200 cue-criterion values and a regression line, for use in the testing phase.

Testing

The purpose of the training phase was to provide subjects with the information and tools necessary to enable them to make optimal predictions for each of the cues alone. In the testing phase, the purpose was to examine how subjects combined information from two sources in order to form a single judgment of the criterion. The combined use of two sources of information was examined in two ways, through the Simultaneous presentation of both cues and through the Successive presentation of the cues.

Simultaneous presentation. In the Simultaneous presentation condition, numbers from both the .80 correlation set and the .40 correlation set were displayed at the same time. The testing phase consisted of 60 trials on which numbers from both cue sets were presented. Over these trials, the correlation between the two sets of cue numbers was approximately zero. Subjects were instructed that both cues provided unique, independent information about the criterion they were to judge. Subjects were allowed

to consult the scatter plots and regression lines at any time. No feedback was provided on any of the 60 testing trials.

Successive presentation. In the Successive presentation condition, a number from either the .80 or the .40 correlation set was first displayed to the subjects (depending on the group to which the subject had been assigned). Subjects were required to predict a criterion number on the basis of this single cue number. Following the subject's first predictions, a second cue number was displayed together with the first number. The subjects were then required to make a second judgment, based on information from both cues.

As in the Simultaneous presentation condition, the independence of the two sources of information was stressed. Subjects were again allowed to consult the scatter plots and regression lines at any time. No feedback was given during the 60 test trials.

Results

Lens model measures. Two of the standard indices of the subjects' performance, r_m and r_s, were computed for each subject. Since in the present task there was no feedback, and thus no actual criterion, r_α was calculated for each subject via Equation 3, using $r_e = .89$. On these measures, the four groups of subjects did not differ significantly (Kruskal-Wallis Analysis of Variance by Ranks), so the results of all groups were combined. Most subjects appeared to perform with reasonable success, the average values of r_α, r_m, and r_s being .70, .99, and .86 respectively.

b-weights. Table 1 summarizes the cue utilizations, i.e., b-weights, computed on the subjects' responses to the joint presentation of two cue values. For the Simultaneous Groups there was no difference in performance between the group that was trained first on the .40 cue and the group that

was trained first on the .80 cue. Therefore, the results from the 20 subjects have been combined. All 20 subjects in the Simultaneous group underweighted the .80 cue, and 18 of them underweighted the .40 cue. The .80 cue was weighted more heavily than the .40 cue by 14 of the 20 subjects.

Insert Table 1 about here

The b-weights computed on the second responses made by the Successive Groups show an order effect: When the .40 cue was presented first, the mean b-weight for it was .18; when it was the second cue, the mean b-weight was .26. This difference was significant (Mann-Whitney U test; $p < .02$). When the .80 cue was presented first, the mean b-weight for it was .32; when the .80 cue was second, its mean b-weight was .49, again significantly higher ($p < .001$). Thus subjects underweighted both cues, but most especially underweighted the first cue they saw.

The low b-weights of the subjects models indicates conservatism. So do the ratios of subjects' b-weights to the optimal weights ($b_{i,s}/b_{i,e}$). For the .40 cue, the mean b-weight ratio across all 40 subjects was .62 (range .23 to 1.23); only two subjects in the Simultaneous Group and one subject in the Successive Group were not conservative (ratio greater than 1) in the use of this cue. For the .80 cue, the mean b-weight ratio was .51 (range .16 to .78); all 40 subjects showed conservatism on this measure.

Revision ratios. For a detailed analysis of revision ratios in the Successive Groups all 60 pairs of stimuli were classified into the following three categories (except for one pair in the .40-.80 condition and nine pairs in the .80-.40 condition for which the optimal second revision was zero; these trials were excluded from the revision analysis):

Category 1: The two cues were on the same side of the mean (50), and the second cue was farther from the mean than the first cue (e.g., 53 followed by 62, or 48 followed by 36). This category is called Increasing.

Category 2: The two cues were on the same side of the mean; the second cue was closer to the mean than the first (e.g., 62 followed by 53, or 36 followed by 48). This category is called Decreasing.

Category 3: The two cues were on opposite sides of the mean (e.g., 62 followed by 48, or 36 followed by 53).

Revision ratios were computed for each subject on each trial. The averages over subjects and over trials within each category are shown in Table 2, along with the number of trials per subject in each category. For the ideal subject, the revision ratio would equal 1.00. These 20 subjects were conservative in their revisions on the Increasing trials. However, on Decreasing trials subjects tended to revise in the wrong direction, i.e., toward the mean. For example, for the .80-.40 Group, the two cues on trial 27 were 64 followed by 54. The optimal responses are, to the first cue, 61, and after both cues, 63, so $d_e = +2$. The average first response across the 10 subjects was 61.2; the average second response was 55.4; the average $d_s = -5.8$. On this trial all 10 subjects responded with a smaller number after both cues than after the first cue. Table 2 also indicates that when the two cues were on different sides of the mean, the subjects, on the average, were counter-conservative, revising too much toward the second cue. The average revision computed across all trials was conservative for both groups.

Insert Table 2 about here

Individual differences. Large individual differences were found in the responses to the first cue made by subjects in the Successive Group. All a subject needed to do to give a perfect performance was to use the appropriate scatter plot (which was in front of him) and read the answer indicated by the regression line, but visual inspection of the scatter of the subjects' responses indicated that not all subjects were "reading off" their responses from the materials available to them. Only about half the subjects clustered their responses close to the regression line. Accordingly, the Successive Group subjects were classified as "Obedient" if 80% or more of their responses to the first cue were within ±2 1/2 units of the optimal response, or "Disobedient" of 65% or fewer of their first responses fell that close to the regression line. In the .40-.80 Group, five subjects were classified as Obedient and five as Disobedient. In the .80-.40 Group, six were Obedient and four Disobedient. The most disobedient subject's first responses correlated only .42 with the cue presented.

The Obedient Successive Groups subjects had slightly, but not significantly, higher performance measures r_{α}, r_m, and r_s. The b-weights computed on their first responses were, of course, all very close to optimal, but their second responses differed not at all from their Disobedient colleagues on the measures shown in Table 1.

The interesting fact about the obedient subjects is the high degree of consistency with which they responded in the directions shown in Table 2. For Increasing trials, 71% of their 167 responses were conservative (i.e., the revision ratio fell between 0 and .99). For Decreasing trials, 78% of their 145 responses were in the wrong direction. The most optimal Obedient subject revised in the wrong direction on 10 of 17 Decreasing

trials. When the two cues were on different sides of the mean, the Obedient subjects were counter-conservative on 63% of their 289 responses. The Disobedient subjects showed the same pattern of revisions but with less consistency, due to the deviant nature of their first responses (which was the defining characteristic of this group).

<u>Search for an averaging strategy</u>. Further analyses were carried out in an attempt to determine whether the performance of the subjects could be accounted for by some sort of averaging strategy. The simplest such strategy, whereby the final response is the average of the two cues, leaves some ambiguity about the response made to the first cue. It seems unlikely that the Obedient subjects, who we know made properly regressed responses to the first cue, later disregarded the regression effect and averaged the two cues to produce their final response. A more reasonable possibility for averaging is that subjects regressed both cues (as they had been trained to do), and responded with the average of the regressed cues. These two averaging strategies may be compared with the optimal strategy, as follows:

$$\text{Optimal response} = b_1\ (X_1 - 50) + b_2\ (X_2 - 50) + 50 \tag{4}$$

$$\text{Averaging response} = 1/2\ (X_1 + X_2) \tag{5}$$

$$\text{Regressed average} = 1/2\ [\overbrace{b_1\ (X_1 - 50) + 50}^{\text{regressed 1st cue}} + \overbrace{b_2\ (X_2 - 50) + 50}^{\text{regressed 2nd cue}}] \tag{6}$$

Only the second averaging strategy (Equation 6), under which the subjects average the regressed values of both cues, predicts the various patterns of revision ratios shown in Table 2. These predictions are illustrated in Table 3, which shows the Obedient subjects' responses to two trials.

Insert Table 3 about here

On trial 30, the .40-.80 Group subjects were first presented with a cue value of 56 followed by a cue value of 60. This trial was coded by the experimenters as an Increasing trial. The .80-.40 Group subjects first saw a cue value of 60, then a cue value of 56, coded as a Decreasing trial. On this trial, the optimal model and the strategy here suggested both call for a first response (which could be read directly off the scatter plots) of about 52 for the .40-.80 Group and 58 for the .80-.40 Group. On the second response the optimal model and the regressed averaging strategy differ. The optimal model calls for a response, by both groups, of about 60, but the proposed strategy calls for a second response, by both groups, of about 55.

The difference, d, between the first response and the second response best illustrates the systematic departure of the regressed averaging strategy from the optimal model. Almost all (93%) of the increasing trials share the characteristic shown by the .40-.80 condition of trial 30 in Table 3: The difference under the optimal model (d_e) is larger than the difference derived from the regressed averaging strategy (d_{ra}). If subjects were following such a strategy, they would appear to be conservative on most increasing trials, as observed.

Most (70%) of the Decreasing trials have the characteristic shown by the .80-.40 condition of trial 30, shown at the upper right in Table 3. The optimal strategy requires that the second response be farther from the mean than the first response, while the regressed averaging strategy yields a second response closer to the mean than the first. Thus if subjects were using the latter strategy, the revision ratio, d_s/d_e, would most often be negative, as observed.

The lower half of Table 3 shows the Obedient subjects' responses to trial 31, on which the two cues were equally spaced on different sides of the mean. The Obedient subjects' responses closely matched the regressed averaging strategy, which yields a conservative revision ratio (-4.8/-6.4) in the .40-.80 condition and a counter-conservative revision ratio (+4.8/+3.2) in the .80-.40 condition. Across all the different-side trials, the Obedient subjects' responses matched the predictions (conservative or counter-conservative) of the regressed averaging strategy 82% of the time.

If subjects were following the regressed averaging strategy rather than the optimal strategy, then the subjects' revisions, d_s, should correlate more highly with revisions yielded by that strategy, d_{ra}, than with the optimal differences, d_e. The two models are sufficiently similar (d_e correlates .87 with d_{ra} in the .40-.80 condition and .47 in the .80-.40 condition) that partial correlations between d_s and d_e, with d_{ra} partialled out, and between d_s and d_{ra}, with d_e partialled out, are more revealing. These correlations, which are summarized in Table 4, indicate substantial relationships between the subjects' responses and the responses derived from the regressed averaging strategy among the Obedient subjects. In contrast, the correlations between subjects' responses and the optimal model are negligible when d_{ra} is partialled out. Even among the Disobedient subjects, whose d_s values were known to be prone to error, all but one subject's responses were predicted much better by the regressed averaging strategy than by the optimal model.

Insert Table 4 about here

Were the subjects in the Simultaneous Groups also using the regressed averaging strategy? Correlational analysis cannot easily discriminate between this strategy and the optimal model, because the predicted responses under both models are perfectly correlated. The mean squared error was computed across the 60 trials for each subject and each model. For the Successive groups the simple averaging model (Equation 5) fit better than the optimal model for 18 of the 20 subjects (with an 18% reduction in squared error). However, the regressed averaging model (Equation 6) even fit better than the simple averaging model for all 20 subjects (with a further 26% reduction in mean squared error). The results were essentially the same for the simultaneous groups. The simple averaging model fit better than the optimal model for all 20 subjects (25% reduction in squared error) while the regressed averaging model fit better than the simple averaging model for 15 out of 20 subjects (19% further reduction in error).

To summarize the analysis of mean squared errors, subjects seemed to be averaging the regressed cues rather than adding their effects as the optimal model does. Whether or not the subjects saw the .40 or the .80 cue first or whether they belonged to the Obedient or Disobedient classification had little or no effect on the superiority of the regressed averaging model over the other contenders.

The regressed averaging strategy can account for the conservative b-weights shown in Table 1 as well as the patterns of conservative, counter-conservative, and wrong direction revision measures shown in Table 2. Among all the reported results, there is only one finding not explained by the regressed averaging strategy: The order or recency effect shown in Table 1 for the Successive subjects. This effect was just as strong

for the Obedient as for the Disobedient subjects, and is not predictable from either the optimal model or the averaging strategy. However, this effect is consistent with findings of other studies of information integration and can be handled by a slightly modified version of the averaging model (see, for example, Anderson, 1968).

Discussion

In the present study, the subjects were taught the relevance of two information sources taken separately, then tested, without feedback, for their ability to integrate several items of information into a single judgment. This particular task is common in Bayesian research, but unique in research with the regression paradigm. According to the usual measures of performance, the subjects in this study appeared to be moderately successful in the task. The three indices of achievement, matching, and linear consistency were similar to those found in previous studies (e.g., Dudycha & Naylor, 1966).

If the analysis of the data had stopped with the global performance measures discussed above, the study might have been presented simply as an illustration of how the regression model and the Bayesian model could be used together, and the results (moderate fit with the regression model and conservatism) could have been viewed as congruent with previous research. But a more molecular examination of the data, departing from the traditional Bayesian and regression analyses, opened new vistas. The hypothesis that the subjects were averaging the regressed values of the two cues was supported by the data. Such behavior is consistent with the averaging strategies discussed by Anderson (1973a, b) and with the representativeness hypothesis proposed by Kahneman and Tversky (1973). This hypothesis implies that the prediction of a criterion should coincide with the most representative

description of the cues. An averaging strategy certainly produces more representative predictions than the optimal model, which sometimes makes predictions that are more extreme than either cue.[3]

In following an averaging strategy, the present subjects were not simply exhibiting a quantitative degree of non-optimality, but were performing in ways qualitatively different from the regression model. The traditional measures derived from this model were not useful in uncovering these serious discrepancies from the models. The results of this study strengthen the belief that molecular analysis of the heuristics that subjects employ is essential to understanding the processes involved in human decision making.

References

Anderson, N. H. Application of a linear-serial model to a personality-impression task wing serial presentation. Journal of Personality and Social Psychology, 1968, 10, 354-362.

Anderson, N. H. Comment on "An analysis-of-variance model for the assessment of configural cue utilization in clinical judgment." Psychological Bulletin, 1969, 72, 63-65.

Anderson, N. H. Algebraic models in perception. In E. C. Carterette and M. P. Friedman (Eds.), Handbook of perception. Vol. 2. New York: Academic Press, 1973. (a)

Anderson, N. H. Information integration theory: A brief survey. In D. H. Krantz, R. C. Atkinson, R. D. Luce, and P. Suppes (Eds.), Contemporary developments in mathematical psychology. 1973, in press. (b)

Beach L. R., Wise, J. A., & Barclay, S. Sample proportion and subjective probability revisions. Organizational Behavior and Human Performance, 1970, 5, 183-190.

Brehmer, B. Cue utilization and cue consistency in multiple-cue probability learning. Organizational Behavior and Human Performance, 1972, 8, 286-296.

Brehmer, B., & Lindberg, L. A. The relation between cue dependency and cue validity in single-cue probability learning with scaled cue and criterion variables. Organizational Behavior and Human Performance, 1970, 5, 542-554.

Brunswik, E. Perception and the representative design of experiments. Berkeley: University of California Press, 1956.

DuCharme, W. M. A response bias explanation of conservative human inference. Journal of Experimental Psychology, 1970, 85, 66-74.

DuCharme, W. M., & Peterson, C. R. Intuitive inference about normally distributed populations. Journal of Experimental Psychology, 1968, 78, 269-275.

Dudycha, L. W., & Naylor, J. C. Characteristics of the human inference process in complex choice behavior situations. Organizational Behavior and Human Performance, 1966, 1, 110-128.

Edwards, W. Conservatism in human information processing. In B. Kleinmuntz (Ed.), Formal representation of human judgment. New York: Wiley, 1968. Pp. 17-52.

Hursch, C., Hammond, K. R., & Hursch, J. L. Some methodological considerations in multiple cue probability studies. Psychological Review, 1964, 71, 42-60.

Kahneman, D., & Tversky, A. Subjective probability: A judgment of representativeness. Cognitive Psychology, 1972, 3, 430-454.

Kahneman, D., & Tversky, A. On the psychology of prediction. Psychological Review, 1973, 80, 237-251.

Naylor, J. C., & Schenck, E. A. ρ_m as an "error-free" index of rater agreement. Educational and Psychological Measurement, 1966, 26, 815-824.

Peterson, C. R., & Beach, L. R. Man as an intuitive statistician. Psychological Bulletin, 1967, 68, 29-46.

Slovic, P. Cue consistency and cue utilization in judgment. American Journal of Psychology, 1966, 79, 427-434.

Slovic, P., & Lichtenstein, S. C. Comparison of Bayesian and regression approaches to the study of information processing in judgment. Organizational Behavior and Human Performance, 1971, 6, 649-744.

Tucker, L. R. A suggested alternative formulation in the development by Hursch, Hammond, & Hursch, and by Hammond, Hursch, & Todd. _Psychological Review_, 1964, 71, 528-530.

Footnotes

1. This research was supported by the Advanced Research Projects Agency of the Department of Defense (ARPA Order No. 2449), and was monitored by ONR under Contract No. N00014-73-C-0438 (NR 197-026). We thank Berndt Brehmer, Robyn Dawes, Lewis Goldberg, Paul Hoffman, Daniel Kahneman, Leonard Rorer, and Amos Tversky for their comments.

2. Requests for reprints should be sent to Sarah Lichtenstein, Oregon Research Institute, P.O. Box 3196, Eugene, Oregon, 97403.

3. Kahneman and Tversky (1973) found that subjects failed to regress when predicting from unreliable single cues. They attributed this to representativeness considerations. The averaging of regressed cues in the present study is undoubtedly due to the heavy emphasis in training on using the regressed regression line. Without such emphasis, it seems likely that subjects would simply have averaged the non-regressed cue values.

Table 1

Subjects' b-Weights After Seeing Both Cues

	$b_{.4,s}$		$b_{.8,s}$		$b_{.8,s}/b_{.4,s}$	
	Mean	Range	Mean	Range	Mean	Range
Simultaneous (n=20)	.27	.10-.49	.40	.16-.60	1.70	.44-3.07
Successive						
.40-.80 (n=10)	.18	.09-.25	.49	.39-.62	3.09	2.00-6.22
.80-.40 (n=10)	.26	.15-.40	.32	.13-.41	1.31	.36-2.00
Optimal Values	.40	-	.80	-	2.00	-

Table 2

Conservatism in Successive Responding:

Mean Revision Ratios[a]

Group	Same Side of Mean		Different Sides of Mean	Mean for All Trials
	Increasing	Decreasing		
.40-.80	0.58	-1.23	1.21	.37
(n[b])	(13)	(17)	(29)	(59)
.80-.40	0.21	-3.17	1.92	.35
(n[b])	(17)	(10)	(24)	(51)

[a]Positive ratios less than unity show conservatism; positive ratios greater than unity show counter-conservatism; negative ratios show revision in the wrong direction.

[b]Number of trials per S in each category.

Table 3

Examples of Individual Subjects' Responses for the Obedient, Successive Group

Trial #30: $Cue_{.40} = 56$ and $Cue_{.80} = 60$

.40-.80 Group				.80-.40 Group			
S#	1st Response	2nd Response	d	S#	1st Response	2nd Response	d
5	52	54	+2	7	58	56	-2
8	52	55	+3	1	58	54	-4
7	52	55	+3	5	58	53	-5
10	50	55	+5	6	58	54	-4
2	52	55	+3	8	58	56	-2
				3	58	56	-2
Mean	51.6	54.8	+3.2	Mean	58.0	54.8	-3.2
Optimal	52.4	60.4	+8.0	Optimal	58.0	60.4	+2.4
Regressed Avg.	52.4	55.2	+2.8	Regressed Avg.	58.0	55.2	-2.8

Trial #31: $Cue_{.40} = 58$ and $Cue_{.80} = 42$

.40-.80 Group				.80-.40 Group			
S#	1st Response	2nd Response	d	S#	1st Response	2nd Response	d
5	52	47	-5	7	44	47	+3
8	53	47	-6	1	44	50	+6
7	53	47	-6	5	41	47	+6
10	50	44	-6	6	43	48	+5
2	53	48	-5	8	43	47	+4
				3	42	48	+6
Mean	52.2	46.6	-5.6	Mean	42.8	47.8	+5.0
Optimal	53.2	46.8	-6.4	Optimal	43.6	46.8	+3.2
Regressed Avg.	53.2	48.4	-4.8	Regressed Avg.	43.6	48.4	+4.8

Table 4

Correlations Based on the Difference Between First Responses and Second Responses for Successive Groups

Correlations Between	.40-.80 Condition Obedient	.40-.80 Condition Disobedient	.80-.40 Condition Obedient	.80-.40 Condition Disobedient
d_e and d_s	.78	.58	.45	.30
d_{ra} and d_s	.92	.63	.85	.65
Partial Correlations Between				
$d_e, d_s \cdot d_{ra}$	-.05	-.04	.14	-.05
$d_{ra}, d_s \cdot d_e$	.78	.42	.81	.62